THE ROYAL SOCIETY

Gene-pool variation in Caledonian and European Scots pine (*Pinus sylvestris* L.) revealed by chloroplast simple-sequence repeats

J. Provan[1], N. Soranzo[1], N. J. Wilson[1], J. W. McNicol[2], G. I. Forrest[3], J. Cottrell[3] and W. Powell[1]*

[1]*Department of Cell and Molecular Genetics, and* [2]*Biomathematics and Statistics Scotland, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK*
[3]*Forest Research, Northern Research Station, Roslin, Midlothian, UK*

We have used polymorphic chloroplast simple-sequence repeats to analyse levels of genetic variation within and between seven native Scottish and eight mainland European populations of Scots pine (*Pinus sylvestris* L.). Diversity levels for the Scottish populations based on haplotype frequency were far in excess of those previously obtained using monoterpenes and isozymes and confirmed lower levels of genetic variation within the derelict population at Glen Falloch. The diversity levels were higher than those reported in similar studies in other *Pinus* species. An analysis of molecular variance (AMOVA) showed that small (3.24–8.81%) but significant ($p \leqslant 0.001$) portions of the variation existed between the populations and that there was no significant difference between the Scottish and the mainland European populations. Evidence of population substructure was found in the Rannoch population, which exhibited two subgroups. Finally, one of the loci studied exhibited an allele distribution uncharacteristic of the stepwise mutation model of evolution of simple-sequence repeats, and sequencing of the PCR products revealed that this was due to a duplication rather than slippage in the repeat region. An examination of the distribution of this mutation suggests that it may have occurred fairly recently in the Wester Ross region or that it may be evidence of a refugial population.

Keywords: Scots pine; *Pinus sylvestris*; chloroplast; simple-sequence repeats

1. INTRODUCTION

The native populations of Scots pine (*Pinus sylvestris* L.) in Scotland represent the only remaining natural pine woodland in the UK (Steven & Carlisle 1959). As a result of excessive exploitation and mismanagement, only a few dozen small woodlands covering less than 11 000 ha ($1\,\text{ha} = 10^4\,\text{m}^2$) remain (Goodier & Bunce 1977). Recently, the perceived historic, scientific and recreational importance of these woodlands has increased and the conservation of native pinewoods has aroused both public and scientific interest. One of the fundamental issues in the conservation of present-day populations of native Scots pine is to maintain high levels of population diversity while protecting the genetic integrity of the native populations. Knowledge of the genetic structure of existing populations should form the basis for a rational and sustainable conservation programme. Current legislation on the management of native Scots pine in Scotland is based on the biochemical analysis of monoterpene composition (Forrest 1980), and this type of analysis has also been used to study the relationships between native and European populations of *P. sylvestris* (Forrest 1982).

The analysis of hypervariable simple-sequence repeats (SSRs), or microsatellites (Tautz 1989; Powell *et al*. 1996*a*), has provided a powerful technique to study genetic diversity and gene flow in natural populations of tree species (Chase *et al*. 1996; Dow & Ashley 1996; Dawson *et al*. 1997; White & Powell 1997*a*,*b*). In conifers, however, the development of SSR markers is not straightforward, as the complexity of conifer genomes makes it difficult to isolate and characterize effective single-locus SSRs (Kostia *et al*. 1995; Morgante *et al*. 1996; Pfeiffer *et al*. 1997). The occurrence of SSR polymorphism has recently been documented in the chloroplasts of pines (Powell *et al*. 1995*a*; Vendramin *et al*. 1996) as well as in soya bean (Powell *et al*. 1995*b*, 1996*b*) and rice (Provan *et al*. 1996, 1997). The combination of data from multiple polymorphic sites in the non-recombining, predominantly uniparentally-inherited haploid chloroplast genome offers a robust, highly informative assay for the analysis of diversity in natural populations of plant species.

We have studied levels of chloroplast SSR (cpSSR) variation in seven native Scottish and eight mainland European populations of Scots pine to quantify levels of

*Author for correspondence (wpowel@scri.sari.ac.uk).

Received 27 April 1998 *Accepted* 1 June 1998

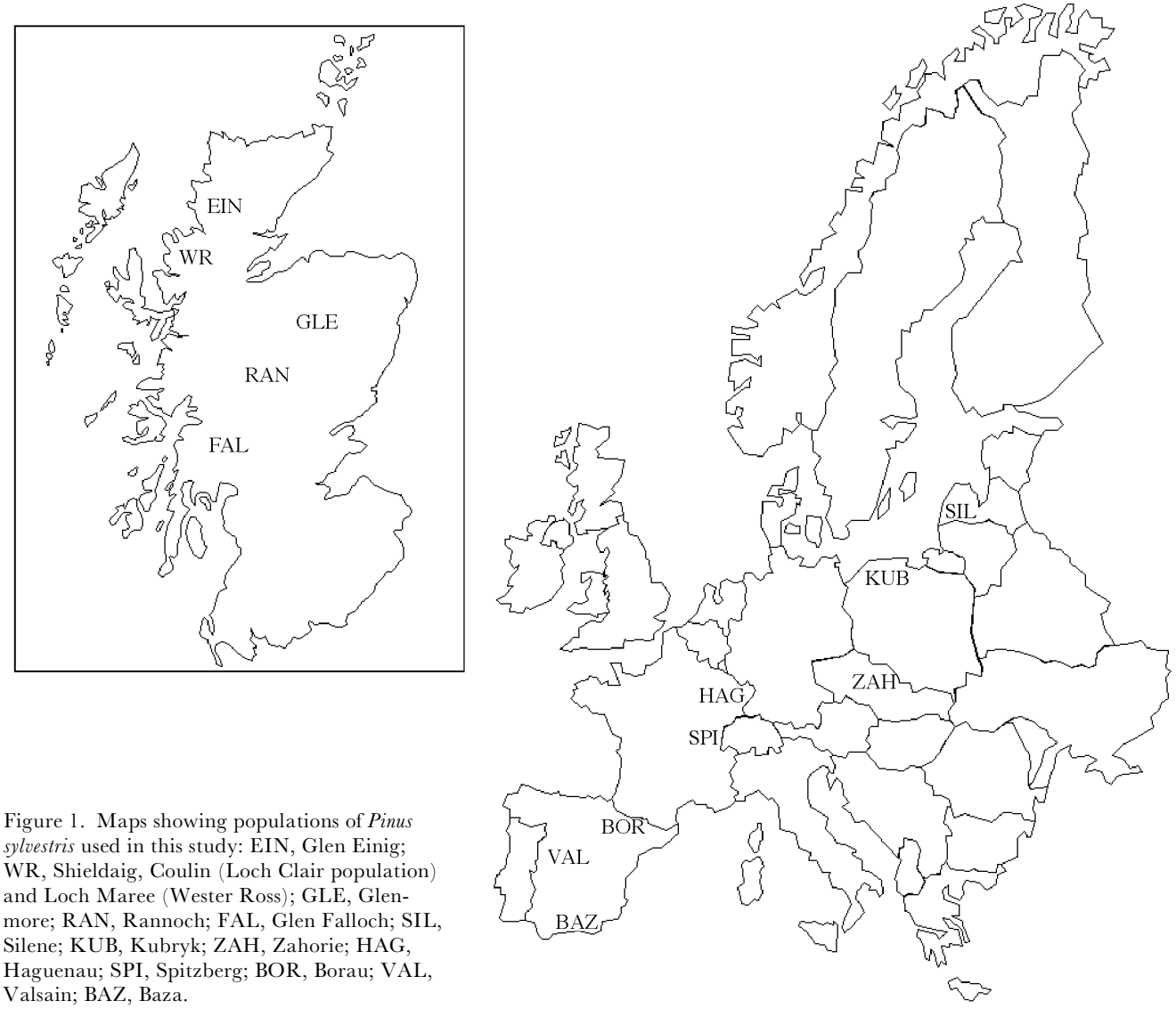

Figure 1. Maps showing populations of *Pinus sylvestris* used in this study: EIN, Glen Einig; WR, Shieldaig, Coulin (Loch Clair population) and Loch Maree (Wester Ross); GLE, Glenmore; RAN, Rannoch; FAL, Glen Falloch; SIL, Silene; KUB, Kubryk; ZAH, Zahorie; HAG, Haguenau; SPI, Spitzberg; BOR, Borau; VAL, Valsain; BAZ, Baza.

differentiation both within and between populations. Our aims were to describe the variation detected and elucidate the mutational process operating at cpSSR regions in *P. sylvestris*, to compare the levels of polymorphism detected using cpSSRs with those obtained previously from monoterpenes and isozyme analyses, and to study the relationships between native Scottish and mainland European populations of Scots pine.

2. MATERIALS AND METHODS

(a) *Genetic material and DNA extraction*

The populations studied are shown in figure 1. Needles were sampled directly from Scottish and Spanish populations. Material from other European populations was obtained from an INRA provenance trial in Orléans National forest. DNA was extracted from needle tissue using the method of Wagner *et al.* (1987).

(b) *Identification of cpSSRs, primer design and polymerase chain reaction (PCR)*

The complete chloroplast sequence of black pine (*Pinus thunbergii*) was analysed for the presence of mononucleotide $(A{\cdot}T)_n$ and $(G{\cdot}C)_n$ repeats using the STRINGSEARCH software (Genetics Computer Group). Nineteen repeats containing $n \geqslant 10$ repeats were identified and primers were designed to amplify 18 mononucleotide SSR repeats using the PRIMER software package (V0.5; table 1). The repeat at the 107 569 base pair (bp) position lay within the coding sequence of the *rpl*32 gene and was not studied because SSR loci within coding regions (excluding the hypervariable trinucleotides associated with various pathological states in mammals) are under greater selective pressure and consequently exhibit little or no polymorphism. PCR was done in a total volume of 10 µl containing 1 × PCR buffer (10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3), 200 µM dNTPs, 10 pmol [^{32}P] end-labelled forward primer (Sambrook *et al.* 1989), 10 pmol reverse primer, 0.1 U *Taq* polymerase (Boehringer Mannheim) and 50 ng of genomic DNA. Reactions were done on an MJ Research PTC 200 DNA engine thermal cycler using the following parameters: (i) initial denaturation at 94 °C for 3 min; (ii) 30 cycles of denaturation at 94 °C for 15 s, annealing at 60 °C for 15 s and extension at 72 °C for 30 s; and (iii) final extension at 72 °C for 5 min. After addition of 10 µl of loading buffer (95% formamide), products were resolved on 6% denaturing polyacrylamide gels containing 1 × TBE buffer and 8 M urea at 80 W constant power for 2 h.

Table 1. Pinus sylvestris *cpSSR primers*

locus	repeat	location	primers (5′–3′)	expected size[a]
PCP1289	$(T)_{17}$	*psb*A/ORF22 intergenic region	TCCTGGTTCCAGAAATGGAG TAATTTGGTTCCAGAATTGCG	113
PCP9434	$(A)_{10}$	ORF59a within *trn*G intron	AAACTGACGTAGATGCCATGG GCGGTATGAGGGAAGAAGC	131
PCP26106	$(A)_{14}$	*rpo*B/ORF46a intergenic region	AATCCGACAAAAAAGATTCGG GCTCCATTTCACGTGGTTG	149
PCP30277	$(A)_{12}(G)_{10}$	ORF46b	TGTTGATGTCGTAGCGGAAG ATGAAATGAATCACTTCCCCC	138
PCP36567	$(T)_{11}$	*psb*J/ORF42b intergenic region	AAAAGAGGAGGAAAAACACCTT AAGAGCAGACAAGTAAGGGGC	115
PCP41131	$(T)_{11}$	ORF119	AAAGCATTTCCAGTTGGGG GGTCAGGATTCATGTTCTTCC	138
PCP45071	$(T)_{15}$	ORF46d/*atp*B intergenic region	ACTGGTCTGATCGACCCAAT TTCTACACTTGCGGAAACCC	149
PCP48256	$(T)_{10}$	ORF64a	ACGTTGGACCAGAGCAGG CGAATTTTTCGAAGAACTAGCG	120
PCP51928	$(T)_{10}$	*trn*S/ORF57b intergenic region	CTTTCTACGGAACGGAAAAGG GCACTGCGGGAAAAAAATAA	141
PCP63771	$(T)_{10}$	*rps*l9/*rpl*2 intergenic region	TGAACGTGCCATGATCAATT GGGGCTATAGTGCACTTGGAA	140
PCP71987	$(T)_{16}$	IRF169 intron 1	TCTTTGCAAGAAGGATGGCT GGGGAGTAATCCGTGGAATT	113
PCP79987	$(A)_{12}$	*trn*S/ORF49b intergenic region	TTTTCAACAATTGCATTTACCG GGCGGGATAGGAGTCTTTTC	120
PCP87314	$(T)_{14}$	*trn*I/*trn*A intergenic region	TCCAGGATAGCCCAGCTG TATATCCCCCGTACTTGGACC	116
PCP100842	$(A)_{12}$	ORF1756/ORF64c intergenic region	TCAATACAAATGATGGGAGTCG TTTTGCCATATCCTGAAACTCC	146
PCP102652	$(T)_{11}$	*jndh*I/ORF43d intergenic region	TTCCCAGATCCATTGAAATACA TATGTGCGCGATAATTTCCA	117
PCP107165	$(A)_{10}$	ORF44d/ORF40f intergenic region	GTTTTGGATCGGAATGGATG CTATCCATTCTGCCTTCCCA	148
PCP109612	$(A)_{11}$	ORF49c/*rps*l2 intergenic region	ATCGAACAACGAGAATAATCCA TTGGGGGTGATAGTGGAAAA	150

[a] From *P. thunbergii* chloroplast sequence.

Gels were transferred onto 3 MM blotting paper (Whatman), dried, and exposed to X-ray film overnight without intensification screens. Amplification of locus PCP110074 failed to give a reproducible product and thus was subsequently excluded from the analysis.

(c) *Direct sequencing of PCR products*

PCR products were sequenced directly using standard fluorescent terminator chemistry on a 373A Applied BioSystems automated sequencer.

(d) *Analysis of cpSSR polymorphism*

Diversity values based on haplotype frequencies were calculated using the equation

$$\hat{H} = n(1 - \Sigma p_i^2)/(n-1), \quad (1)$$

where n is the number of individuals analysed and p_i is the frequency of the ith haplotype. The sampling variance was calculated as

$$V(\hat{H}) = 2[2(n-2)[\sigma p_i^3 - (\Sigma p_i^2)^2] + \Sigma p_i^2 - (\Sigma p_i^2)^2], \quad (2)$$

where p_i is the frequency of the ith haplotype (Nei 1987). Principal coordinate analysis was done based on city-block similarities (Digby & Kempton 1989) using the GENSTAT software package (version 5.0; 1987). Levels of population differentiation were estimated from Φ-statistics (Excoffier *et al*. 1992; Michalakis & Excoffier 1996) and were calculated using the Arlequin software (version 1.1; Schneider *et al*. 1997). Distances between cpSSR haplotypes were calculated from the sum of the squared number of repeat differences between two haplotypes:

$$\hat{d}_{xy} = \Sigma(a_{x_i} - a_{y_i})^2, \quad (3)$$

where a_{x_i} and a_{y_i} are the number of repeats for the ith locus in haplotypes x and y. This gives an analogue of Slatkin's R_{ST} (1995) for population differentiation.

3. RESULTS

(a) *Patterns of genetic variation detected at cpSSR regions*

Thirteen of the 17 cpSSR loci analysed were polymorphic, exhibiting between two and six alleles (table 2*a*,*b*). From these allele frequencies it can be seen that most loci exhibited a unimodal distribution, with alleles differing by 1 bp from each other. This suggests that most of the observed variation conforms to the stepwise

Table 2 *Allele distribution at polymorphic cpSSR loci in populations of* P. sylvestris
(The most common allele at each locus in each population is shown in bold.)
(*a*) Allele distribution at five polymorphic cpSSR loci in seven native and eight mainland European populations of *P. sylvestris*

locus	allele (bp)	population FAL	RAN	COU	SHI	EIN	MAR	GLE	ZAH	KUB	SPI	SIL	HAG	BAZ	BOR	VAL
PCP26106	145	—	0.085	0.085	0.021	0.149	0.170	0.106	0.222	0.238	0.130	0.059	0.211	0.087	0.043	0.042
	146	**0.979**	**0.851**	**0.745**	**0.872**	**0.787**	**0.766**	**0.872**	**0.778**	**0.762**	**0.870**	**0.882**	**0.789**	**0.913**	**0.883**	**0.958**
	147	0.021	0.064	0.170	0.106	0.064	0.064	0.021	—	—	—	0.059	—	—	0.174	—
PCP30277	131	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.125
	132	0.042	0.043	0.021	0.021	0.043	0.128	0.064	0.056	—	0.043	0.176	0.158	0.087	0.130	0.083
	133	0.313	0.149	0.191	0.106	0.149	0.149	0.298	0.333	0.143	0.087	**0.588**	0.158	0.130	0.174	0.250
	134	**0.417**	**0.383**	**0.404**	0.213	**0.489**	**0.468**	**0.383**	**0.553**	**0.333**	**0.609**	0.235	**0.579**	0.043	**0.305**	**0.333**
	135	0.146	0.170	0.170	**0.340**	0.170	0.064	0.106	0.056	**0.333**	0.087	—	0.105	**0.523**	0.130	0.167
	136	0.063	0.255	0.191	0.128	0.149	0.191	0.128	—	0.190	0.130	—	—	0.174	0.218	0.042
	137	0.021	—	0.021	0.191	—	—	0.021	—	—	0.043	—	—	0.043	0.043	—
PCP45071	147	—	—	—	—	—	—	—	—	—	—	0.059	—	—	—	—
	148	—	—	—	—	—	0.021	0.043	—	—	—	—	—	—	—	—
	149	0.125	0.170	0.149	0.170	0.128	0.213	0.298	0.167	0.095	0.174	0.353	0.053	0.087	0.174	0.125
	150	**0.833**	**0.787**	**0.809**	**0.723**	**0.766**	**0.660**	**0.511**	**0.833**	**0.810**	**0.739**	**0.529**	**0.894**	**0.913**	**0.739**	**0.750**
	151	0.042	—	0.021	0.085	0.085	0.064	0.149	—	0.095	0.087	0.059	0.053	—	0.087	0.125
	152	—	0.043	0.021	0.021	0.021	0.043	—	—	—	—	—	—	—	—	—
PCP71987	107	—	—	—	—	—	—	—	—	—	—	0.059	—	—	—	—
	108	—	0.021	—	—	—	—	—	—	—	—	0.059	—	—	0.043	—
	109	0.125	0.128	—	0.106	0.340	0.064	0.043	0.222	0.095	0.043	0.118	0.105	0.391	0.087	0.292
	110	0.271	**0.468**	**0.553**	**0.532**	**0.511**	**0.574**	**0.660**	**0.722**	**0.714**	**0.565**	0.294	**0.474**	**0.566**	**0.696**	**0.500**
	111	**0.479**	0.298	0.426	0.277	0.149	0.362	0.277	0.056	0.191	0.392	**0.412**	0.263	0.043	0.174	0.208
	112	0.104	0.064	—	0.085	—	—	0.021	—	—	—	0.059	0.158	—	—	—
	113	0.021	0.021	0.021	—	—	—	—	—	—	—	—	—	—	—	—
PCP87314	114	—	0.064	—	0.021	0.021	—	—	—	0.095	0.087	—	0.105	—	0.087	—
	115	**0.938**	**0.766**	**0.787**	**0.787**	**0.915**	**0.894**	**0.681**	**0.833**	**0.857**	**0.740**	**0.882**	**0.842**	**0.739**	**0.783**	**0.750**
	116	0.063	0.149	0.191	0.149	0.064	0.085	0.213	0.111	0.048	0.087	0.118	0.053	0.261	0.130	0.167
	117	—	—	0.021	0.043	—	0.021	0.106	0.056	—	0.043	—	—	—	—	0.083
	118	—	0.021	—	—	—	—	—	—	—	0.043	—	—	—	—	—

(*b*) Allele distribution at eight polymorphic cpSSR loci in seven native *P. sylvestris* populations

locus	allele (bp)	population FAL	RAN	COU	SHI	EIN	MAR	GLE
PCP1289	105	—	—	—	—	—	0.021	—
	106	0.104	0.106	0.128	0.085	0.170	0.128	0.043
	107	**0.771**	**0.851**	**0.787**	**0.787**	**0.766**	**0.809**	**0.894**
	108	0.125	0.043	0.064	0.106	0.043	0.043	0.064
	109	—	—	0.021	0.021	0.021	0.021	—
PCP36567	106	0.083	0.213	0.149	0.064	0.085	0.277	0.191
	107	**0.896**	**0.787**	**0.809**	**0.830**	**0.915**	**0.660**	**0.766**
	116	0.021	—	0.043	0.106	—	0.064	0.043
PCP41131	135	—	—	0.064	—	—	—	—
	136	—	—	—	—	—	—	0.064
	137	**0.958**	**1.000**	**0.830**	**0.936**	0.894	**0.915**	**0.872**
	138	0.042	—	0.106	0.064	0.106	0.085	0.064
PCP48256	118	—	0.021	—	0.043	—	—	—
	119	**0.958**	**0.957**	**0.979**	**0.957**	**1.000**	**0.979**	**1.000**
	120	0.042	0.021	0.021	—	—	0.021	—
PCP63771	137	**0.938**	**1.000**	**1.000**	**0.979**	**0.979**	**0.979**	**1.000**
	138	0.062	—	—	0.021	0.021	0.021	—
PCP71987	115	0.042	—	—	—	0.043	0.021	—
	116	**0.875**	**1.000**	**1.000**	**1.000**	**0.957**	**0.979**	**1.000**
	117	0.083	—	—	—	—	—	—
PCP100842	143	—	0.043	—	—	0.021	0.021	—
	144	**1.000**	**0.957**	**1.000**	**1.000**	**0.979**	**0.979**	**1.000**
PCP102652	112	—	—	—	0.021	—	—	—
	113	**1.000**	**1.000**	**0.979**	**0.979**	**0.979**	**1.000**	**0.979**
	114	—	—	0.021	—	0.021	—	0.021

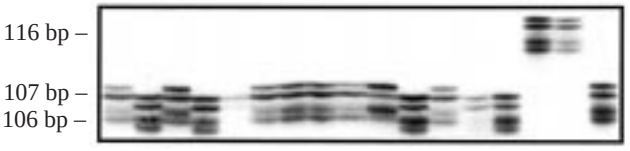

Figure 2. Example of allelic variation detected at locus PCP36567. Allele sizes are based on the size of the 'upper product'. Lower bands are 'stutter' bands commonly found in PCR amplification of SSR regions.

mutation model (SMM), which can be explained by replication slippage at SSR loci (Valdes *et al*. 1993; DiRienzo *et al*. 1994; Freimer & Slatkin 1996). Furthermore, the non-recombinational nature of the chloroplast genome means that unequal crossing-over cannot be the main causal mechanism. It is therefore feasible to use a metric to estimate genetic distance that is based on the SMM and takes into account the cpDNA molecule as a single heritable unit (Morgante *et al*. 1997). This measure of genetic distance is a modification of the $\delta\mu^2$ metric of Goldstein *et al*. (1995). One exception was locus PCP36567, which exhibited a bimodal allele distribution inconsistent with the SMM and was thus excluded from analyses of population differentiation.

(b) *Unusual genotypes at locus PCP36567*

Analysis of the distribution of alleles at each locus generally exhibited an approximately symmetric, unimodal distribution typically associated with the bidirectional mutation processes operating at SSR loci. At locus PCP36567, however, individuals were found that possessed an allele (116 bp) much larger than the others (106/107 bp; figure 2). Because this type of allelic distribution is uncharacteristic of a single SMM of SSR evolution, both the 107-bp and the 116-bp products were sequenced. It was found that the presence of the 116-bp allele was due to the duplication of a 9-bp stretch upstream of the SSR region, rather than a 9-bp increase in the SSR itself (figure 3). Analysis of the distribution of this mutation is shown in table 3, and it can be seen that it was only found in five out of the seven Scottish populations and was completely absent from the mainland European populations studied.

(c) *Levels of genetic variation within populations of* P. sylvestris

Using 13 polymorphic loci, 174 different haplotypes were found in the 330 individuals from the Scottish populations, of which 124 (72%) were unique, i.e. were only found in a single individual (table 4). This means that 38% (124 out of 330) of the trees studied could be uniquely genotyped. Between 47.1% and 80.0% of the haplotypes found in each population were unique to that particular population sample, and between 38.3% and 63.8% of individuals within each sample contained private haplotypes, i.e. those found only in a single population. Over the whole study, 48.8% of individual trees contained a private haplotype. Effective numbers of haplotypes in each population ranged from 14.286 in Glen Falloch to 29.412 in Glenmore. Intra-population gene diversity values based on haplotype frequency ranged from 0.950 in Glen Falloch to 0.987 in Glen Einig. The total haplotype diversity value calculated across all populations was 0.991.

Six of the primer pairs (PCP26106, PCP3027, PCP36567, PCP45071, PCP71987 and PCP87314) were also used to analyse the eight European populations. Because locus PCP36567 did not appear to conform to a single stepwise mutational process, it was excluded from population analyses. Using the other five primers, 133 haplotypes were identified in 515 individuals. Effective numbers of haplotypes based on five loci ranged from 8.180 (Baza) to 25.000 (Rannoch). Levels of gene diversity are shown in table 5 and ranged from 0.908 to 0.981 within individual populations. The total gene diversity across all populations was 0.977. Such diversity levels are much higher than those previously reported in cpSSR studies in other *Pinus* species (Powell *et al*. 1995*a*; Morgante *et al*. 1997). Values for the percentage of individuals within a population that exhibited a private haplotype ranged from 0% in the Zahorie population, which contained no private haplotypes, to 35.29% in Silene.

(d) *Inter- and intra-population differentiation in* P. sylvestris

An analysis of molecular variance (AMOVA) was used to partition the cpSSR variation into within- and between-woodland components (table 6). Considering the data generated with 17 cpSSRs on the Scottish populations, most of the variation (96.76%) was detected within woodlands. A small (3.24%) but significant ($p=0.00101$) portion of the variation was attributable to variation between populations. When the same analysis was extended to examine the 15 populations (Scotland and mainland Europe) based on five cpSSR loci, it was found that 8.81% of the variation ($p<0.00001$) was due to differences between the populations. Partitioning the variation between Scottish and mainland European populations indicated that a small (1.48%), non-significant ($p=0.13306$) portion of the total variation existed between these two groups. This analysis showed that 8.03% of the variation ($p<0.00001$) existed as variation between populations within the two groups, but that most of the variation was attributable to the within-population component (90.49%; $p<0.00001$).

Principal coordinate analysis based on city-block similarities was used to assess levels of intra-population genetic structuring within each woodland. In general, the accessions within a woodland formed a single major cluster. The exception to this was the Rannoch population, which was clearly divided into two subgroups (figure 4). An AMOVA based on these two subpopulations showed that 13.28% ($p=0.00880$) of the variation within the whole population existed between the two subgroups (data not shown).

4. DISCUSSION

Diversity values based on haplotype frequency in the seven native Scottish populations of Scots pine studied (0.950–0.987) were much higher than the earlier monoterpene (0.272–0.378) and isozyme (0.291–0.311) analyses on the same populations (Kinloch *et al*. 1986). These results, along with the high percentage of unique haplotypes (72%) identified in the samples studied and the fact

```
P. thunbergii          AAAAACACCT TTTATTCATT .........T TCTCTCATTA
P. sylvestris (107 bp) AAAAACACCT TTTATTCATT .........T TCTCTCATTA
P. sylvestris (116 bp) AAAAACACCT TTTATTCATT TTCTTCATTT TCTCTCATTA

P. thunbergii          GGAACAAACC CTATCAAAAG TTTTATAGGT TTTTTTTTTT
P. sylvestris (107 bp) GGAACAAACC CTATCAAAAG TTTTATAGGT TTTTTT....
P. sylvestris (116 bp) GGAACAAACC CTATCAAAAG TTTTATAGGT TTTTTT....

P. thunbergii          ATGAGTGATA TTGCCCCTTA CTTGTCTGCT CTTCCTTCT
P. sylvestris (107 bp) ATGAGTGATA TTGCCCCTTA CTTGTCTGCT CTT
P. sylvestris (116 bp) ATGAGTGATA TTGCCCCTTA CTTGTCTGCT CTTA
```

Figure 3. DNA sequence alignment at locus PCP36567. The SSR region is shown in bold and the 9-bp duplication in the 116-bp allele is underlined.

Table 3. *Distribution of the 116-bp allele at locus PCP36567 in 330 individuals from seven Scottish populations*

	116-bp allele	
population	number	% total occurrence
Shieldaig	5	38.47
Loch Maree	3	23.08
Coulin	2	15.38
Glenmore	2	15.38
Glen Falloch	1	7.69

Table 4. *Diversity statistics for Scottish* P. sylvestris *populations based on 17 polymorphic cpSSR loci*

(n=number of individuals sampled; N=number of haplotypes; N_e=effective number of haplotypes: $1/\Sigma p_i^2$; H=gene diversity, $(n(1-\Sigma p_i^2))/(n-1)$; p_i=frequency of the ith haplotype.)

population	n	N	N_e	H	% individuals with private haplotypes
Coulin	47	35	27.778	0.985±0.007	44.71
Glen Einig	47	35	19.231	0.969±0.017	55.28
Glen Falloch	48	30	14.286	0.950±0.020	43.83
Glenmore	47	37	29.412	0.987±0.008	63.78
Loch Maree	47	36	28.571	0.986±0.008	38.32
Rannoch	47	35	28.571	0.986±0.007	40.40
Shieldaig	47	33	27.027	0.984±0.007	59.62
total	330	174	83.312	0.991±0.003	48.84

that 38% of trees could be individually genotyped, demonstrate the power of cpSSR haplotypes for uniquely differentiating individual trees.

The distribution of cpSSR haplotypes over the seven Scottish populations suggests that, at least at the haplotype level, cpSSRs offer potential for the identification of population-specific markers. Between 30 and 37 haplotypes were found in each population and between 47.1% and 80.0% of these were unique to that population (private haplotypes). In terms of individual trees within the woodlands, between 38.3% and 63.8% (average 48.8%) contained a private haplotype. Many of these, however, were only present in a single tree, which suggests that their occurrence in a single population is a result of the relatively limited sample sizes and that they may not be true population markers.

The lowest diversity value and effective number of haplotypes (0.950 and 14.286, respectively) were both found in the Glen Falloch population. These figures reflect a severe depletion of the genetic base of the population. This is due to a great reduction in the population size over the past 200 years to the point where less than 100 trees remain today, resulting in high levels of inbreeding. Both Boyle & Malcolm (1985) and Faulkner (1977) have documented the need for conservation of this derelict population and the high resolving power of the cpSSR technique will be useful in the identification of individuals for the establishment of a clonal breeding orchard with maximum genetic diversity.

Although the mutation rate at cpSSR loci is higher than substitution rates in the rest of the chloroplast genome, it is unlikely that it is of the order of magnitude of the migration rate. Our previous studies on the narrow endemic *P. torreyana* showed no variation at the same 17 cpSSRs in two reproductively isolated populations which underwent a severe bottleneck before their separation around 8000 Ma BP (Ledig & Conkle 1983; Waters & Schaal 1991). Based on these results, the mutation rates at cpSSR loci would not appear to occur at a level that would confound analyses of population genetic structure. This has been confirmed by other cpSSR-based population studies in pines (Morgante *et al.* 1997; Echt *et al.* 1998).

The Φ-statistics calculated from hierarchical analysis based on cpSSR haplotypes indicate fairly low levels of population differentiation, with between 3.24% and 8.81% of the total observed variation existing between populations and no significant differences between Scottish and mainland European populations. This is to

Table 5. *Diversity statistics for Scottish and European* P. sylvestris *populations based on five polymorphic cpSSR loci*

(n=number of indviduals sampled; N=number of haplotypes; N_e=effective number of haplotypes, $1/\Sigma p_i^2$; H=gene diversity, $(n(1-\Sigma p_i^2))/(n-1)$; p_i=frequency of the ith haplotype.)

population	n	N	N_e	H	% individuals with private haplotypes
Coulin	47	27	20.833	0.972±0.009	10.64
Glen Einig	47	26	12.987	0.942±0.023	14.89
Glen Falloch	48	22	9.901	0.918±0.023	10.42
Glenmore	47	31	20.000	0.970±0.013	14.89
Loch Maree	47	31	23.256	0.978±0.009	31.91
Rannoch	47	31	25.000	0.981±0.008	12.77
Shieldaig	47	29	22.727	0.976±0.009	25.53
total (Scotland)	330	101	35.772	0.975±0.007	17.293
Baza	30	10	8.180	0.908±0.027	10.00
Borau	28	21	14.520	0.965±0.023	25.00
Haguenau	19	15	10.989	0.959±0.036	31.58
Kubryk	21	16	14.286	0.976±0.020	19.05
Silene	17	14	10.000	0.956±0.044	35.29
Spitzberg	23	16	12.346	0.961±0.026	17.39
Valsain	29	18	14.750	0.965±0.016	26.00
Zahorie	18	12	10.101	0.954±0.030	0.00
total (Europe)	185	72	25.780	0.967±0.014	20.539
total	515	133	40.175	0.977±0.005	19.024

Table 6. *Analysis of molecular variance (AMOVA) based on cpSSRs*

(Inter-haplotype distances were calculated from sum of squared allele differences ($\delta\mu^2$). (*a*) Scottish populations genotyped with 16 cpSSRs. (*b*) Scottish and mainland European populations genotyped with a subset of five cpSSRs. p values were calculated by non-parametric permutation procedure.)

source of variation	d.f.	variance component	% variation	p
(*a*) between Scottish populations	6	0.1758	3.24	$p=0.0010$
within populations	324	5.2478	96.76	
(*b*) between all populations	14	0.2526	8.81	$p<0.0001$
Scotland versus mainland Europe	1	0.0428	1.48	$p=0.1331$
between populations within groups	13	0.2319	8.03	$p<0.0001$
within populations	501	2.6136	90.49	

be expected for conifers, as they are highly heterozygous outbreeders. Higher values have been reported (e.g. Szmidt *et al*. (1996) in *P. merkusii*), but these are generally due to biogeographical factors such as high levels of fragmentation. Previous studies on *P. sylvestris* (Wang *et al*. 1991; Goncharenko *et al*. 1994; Shigapov *et al*. 1995) have shown that inter-population variation accounted for between 1.7% and 7.5% of the total observed variation. Owing to the pollen-transmitted nature of chloroplast-based markers, they may be expected to show lower levels of population differentiation than mitochondrial markers (Dong & Wagner 1994; Ennos 1994; Latta & Mitton 1997), although high levels of population differentiation have been reported in cpSSR-based studies of *P. leucodermis* (Powell *et al*. 1995*a*) and *P. halepensis* (Morgante *et al*. 1997). The seed-specific nature of mitochondrial markers offers more scope for differentiating between populations and a recent study of mtDNA variation in Scottish *P. sylvestris* populations recorded a value of $F_{ST}=0.370$ (Sinclair *et al*. 1997). Other studies employing such markers have yielded estimates of G_{ST}/F_{ST} to between 0.560 and 0.958 (Dong & Wagner 1993; Strauss *et al*. 1993; Latta & Mitton 1997).

The apparent existence of two genetically different subgroups within a population was a phenomenon found only in the Rannoch population. It is possible that these two subgroups represent two original progenitors or seed lots, as Anderson (1967) has reported that plantations were often located in close proximity to native stands and that gene flow could occur between the two. Alternatively, it has been suggested that the Rannoch woodland was replanted in the past few hundred years using material from a different population (C. Fleming, personal communication) or that biased seed collection from a restricted number of trees was used as a source of planting material. Any of these scenarios could give rise to this apparent genetic substructuring, although the fairly low levels of between-population diversity may preclude such a pronounced effect.

Because SSR loci undergo bidirectional mutation, allelic distribution at these loci is generally modal. At locus PCP36567, however, we detected an unusual rare allele (116 bp) in five of the native Scottish populations; sequencing of the PCR products revealed that the non-uniform allele distribution was due to a 9-bp duplication in the 116-bp allele rather than length polymorphism in the repeat region. This allele was not found in any of the mainland European populations and an examination of the occurrence of the 116-bp allele in the Scottish populations (table 3) revealed that 10 out of the 13 individuals (77%) that exhibited this allele belonged to populations from the Wester Ross region (Shieldaig, Coulin and Loch Maree). This may suggest a fairly recent mutation that occurred in this region and spread slowly through Scotland before fragmentation. Alternatively, this variant may be indicative of the presence of a refugial population in the west of Scotland. Sinclair *et al*. (1997) reported the occurrence of a localized mtDNA haplotype in western Scottish populations and these results are comparable with isozyme and monoterpene data (Kinloch *et al*. 1986), which found that Shieldaig was distinct from the remaining populations.

We have demonstrated here that cpSSR polymorphism can be used to analyse diversity in the chloroplast genome of *P. sylvestris*. In addition, multiple cpSSR loci can be analysed to give haplotypes that may have value as population-specific markers. Ultimately, when cpSSR polymorphisms are considered along with nuclear and mitochondrial markers to give an indication of the overall history and genetic architecture of our native pinewoods,

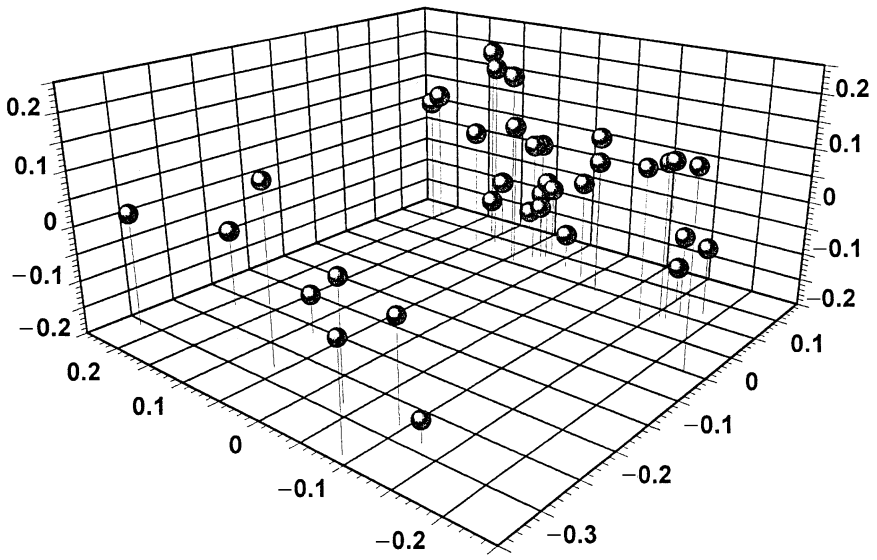

Figure 4. Three-dimensional principal coordinate analysis showing genetic substructuring within the Rannoch population.

this understanding will be an invaluable asset when considering legislative decisions concerning the management of existing populations of Scots pine.

This project is funded by the Scottish Office Agriculture, Environment and Fisheries Department. We thank Charlie Fleming and Bill Brown of the Forestry Commission for assistance and advice on sampling, and Dr C. J. A. Samuel for organizing the collection of material from mainland European populations.

REFERENCES

Anderson, M. L. 1967 *A history of Scottish forestry*. London: Nelson.

Boyle, T. J. B. & Malcolm, D. C. 1985 The reproductive potential and conservation value of a near-derelict Scots pine remnant in Glen Falloch. *Scot. Forestry* **39**, 288–302.

Chase, M., Kessell, R. & Bawa, K. 1996 Microsatellite markers for population and conservation genetics of tropical trees. *Am. J. Bot.* **83**, 51–57.

Dawson, I. K., Waugh, R., Simons, A. J. & Powell, W. 1997 Simple sequence repeats provide a direct estimate of pollen-mediated gene dispersal in the tropical tree *Gliricidia sepium*. *Molec. Ecol.* **6**, 179–183.

Digby, P. G. N. & Kempton, R. A. 1989 *Multivariate analysis of ecological communities*, pp. 19–22. London: Chapman & Hall.

DiRienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M. & Freimer, N. B. 1994 Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natn. Acad. Sci. USA* **91**, 3166–3170.

Dong, J. & Wagner, D. B. 1993 Taxonomic and population differentiation of mitochondrial diversity in *Pinus banksiana* and *Pinus contorta*. *Theor. Appl. Genet.* **86**, 573–578.

Dong, J. & Wagner, D. B. 1994 Paternally inherited chloroplast polymorphism in *Pinus*: estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics* **136**, 1187–1194.

Dow, B. D. & Ashley, M. V. 1996 Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Molec. Ecol.* **5**, 615–627.

Echt, C. S., De Verno, L. L., Anzeide, M. & Vendramin, G. G. 1998 Chloroplast microsatellites reveal population genetic diversity in red pine *Pinum resinosa* Ait. *Molec. Ecol.* **7**, 307–316.

Ennos, R. A. 1994 Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**, 250–259.

Excoffier, L., Smouse, P. E. & Quattro, J. M. 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Faulkner, R. 1977 The gene-pool of Caledonian Scots pine—its conservation and uses. In *Native pinewoods of Scotland: proceedings of the Aviemore Symposium* (ed. R. G. Bunce & A. Jeffers), pp. 96–99. Cambridge: Institute of Terrestrial Ecology.

Forrest, G. I. 1980 Genotypic variation among native Scots pine populations in Scotland based on monoterpene analysis. *Forestry* **53**, 101–128.

Forrest, G. I. 1982 Relationship of some European Scots pine populations to native Scottish woodlands based on monoterpene analysis. *Forestry* **55**, 109–137.

Freimer, N. B. & Slatkin, M. 1996 Microsatellites: evolution and mutational processes. In *Variation in the human genome:* Ciba Foundation Symposium 197, pp. 51–72. London: Ciba Foundation.

Genstat 5.0 Committee 1987 *GENSTAT 5.0 reference manual*. Oxford: Clarendon Press.

Goldstein, D. B., Ruiz-Linares, A., Cavalli-Sforza, L. L. & Feldman, M. W. 1995 Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natn. Acad. Sci. USA* **92**, 6723–6727.

Goncharenko, G. G., Silin, A. E. & Padutov, V. E. 1994 Allozyme variation in natural populations of Eurasian pines. 3. Population structure, diversity, differentiation and gene flow in central and isolated populations of *Pinus sylvestris* L. in eastern Europe and Siberia. *Silvae Genetica* **43**, 119–132.

Goodier, R. & Bunce, R. G. H. 1977 The native pinewoods of Scotland: the current state of the resource. In *Native pinewoods of Scotland: proceedings of the Aviemore Symposium* (ed. R. G. Bunce & A. Jeffers), pp. 78–87. Cambridge: Institute of Terrestrial Ecology.

Kinloch, B. B., Westfall, R. D. & Forrest, G. I. 1986 Caledonian Scots pine: origins and genetic structure. *New Phytol.* **104**, 703–729.

Kostia, S., Varvio, S.-L., Vakkari, P. & Pulkinnen, P. 1995 Microsatellite sequences in a conifer *Pinus sylvestris*. *Genome* **38**, 1244–1248.

Latta, R. G. & Mitton, J. B. 1997 A comparison of population differentiation across four classes of gene marker in limber pine (*Pinus flexilis* James). *Genetics* **146**, 1153–1163.

Ledig, F. T. & Conkle, M. R. 1983 Gene diversity and genetic structure in a narrow endemic torrey pine (*Pinus torreyana* Parry ex. Carr). *Evolution* **37**, 79–85.

Michalakis, Y. & Excoffier, L. 1996 A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**, 1061–1064.

Morgante, M., Pfeiffer, A., Costacurta, A. & Olivieri, A. M. 1996 Molecular tools for population and ecological genetics in coniferous trees. *Phyton-annales Rei Botanicae* **36**, 129–138.

Morgante, M., Felice, N. & Vendramin, G. G. 1997 Analysis of hypervariable chloroplast microsatellites in *Pinus halepensis* reveals a dramatic genetic bottleneck. In *Molecular tools for screening biodiversity: plants and animals* (ed. A. Karp, P. G. Issac & D. S. Ingrams), pp. 407–412. London: Chapman & Hall.

Nei, M. 1987 *Molecular evolutionary genetics*. New York: Columbia University Press.

Pfeiffer, A., Olivieri, A. M. & Morgante, M. 1997 Identification and characterisation of microsatellites in Norway spruce (*Picea abies* K.). *Genome* **40**, 411–419.

Powell, W., Morgante, M., McDevitt, R., Vendramin, G. G. & Rafalski, J. A. 1995*a* Polymorphic simple sequence repeat regions in the chloroplast genome: applications to the population genetics of pines. *Proc. Natn. Acad. Sci. USA* **92**, 7759–7763.

Powell, W., Morgante, M., Andre, C., McNicol, J. W., Machray, G. C., Doyle, J. J., Tingey, S. V. & Rafalski, J. A. 1995*b* Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr. Biol*. **5**, 1023–1029.

Powell, W., Machray, G. C. & Provan, J. 1996*a* Polymorphism revealed by simple sequence repeats. *Trends Pl. Sci*. **1**, 215–222.

Powell, W., Morgante, M., Doyle, J. J., McNicol, J. W., Tingey, S. V. & Rafalski, J. A. 1996*b* Genepool variation in genus *Glycine* subgenus *Soja* revealed by polymorphic nuclear and chloroplast microsatellites. *Genetics* **144**, 793–803.

Provan, J., Corbett, G., Waugh, R., McNicol, J. W., Morgante, M. & Powell, W. 1996 DNA fingerprints of rice (*Oryza sativa*) obtained from hypervariable chloroplast simple sequence repeats. *Proc. R. Soc. Lond*. B **263**, 1275–1281.

Provan, J., Corbett, G., McNicol, J. W. & Powell, W. 1997 Chloroplast DNA variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. *Genome* **40**, 104–110.

Sambrook, J., Fritch, E. F. & Maniatis, T. 1989 *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press.

Schneider, S., Kueffer, J.-M., Roessli, D. & Excoffier, L. 1997 Arlequin v. 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Shigapov, Z. K., Bakhtiyarova, P. M. & Yanbaev, Y. A. 1995 Genetic variation and differentiation in natural populations of the common pine *Pinus sylvestris* L. *Genetika* **31**, 1386–1393.

Sinclair, W. T., Moncur, J. D. & Ennos, R. A. 1997 Multiple origins for Scots pine (*Pinus sylvestris* L.) in Scotland: evidence from mitochondrial DNA variation. *Heredity* **80**, 233–240.

Slatkin, M. 1995 A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462.

Steven, H. M. & Carlisle, A. 1959 *The native pinewoods of Scotland*. London: Oliver & Boyd.

Strauss, S. H., Hong, Y.-P. & Hipkins, V. D. 1993 High levels of population differentiation for mitochondrial DNA haplotypes in *Pinus radiata*, *muricata* and *attenuata*. *Theor. Appl. Genet*. **86**, 605–611.

Szmidt, A. E., Wang, X. R. & Changtragoon, S. 1996 Contrasting patterns of genetic diversity in 2 tropical pines—*Pinus kesiya* Royle ex Gordon and *Pinus merkusii* Jungh et DeVriese. *Theor. Appl. Genet*. **92**, 436–444.

Tautz, D. 1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucl. Acids Res*. **17**, 6463–6471.

Valdes, A. M., Slatkin, M. & Freimer, N. B. 1993 Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133**, 737–749.

Vendramin, G. G., Lelli, L., Rossi, P. & Morgante, M. 1996 A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Molec. Ecol*. **5**, 595–598.

Wagner, D. B., Furnier, G. R., Saghai-Maroof, M. A., Williams, S. M., Dancik, B. P. & Allard, R. W. 1987 Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proc. Natn. Acad. Sci. USA* **84**, 2097–2100.

Wang, X. R., Szmidt, A. E. & Lindgren, D. 1991 Allozyme differentiation among populations of *Pinus sylvestris* (L.) from Sweden and China. *Hereditas* **114**, 219–226.

Waters, E. R. & Schaal, B. A. 1991 No variation is detected in the chloroplast genome of *Pinus torreyana*. *Can. J. For. Res*. **21**, 1832–1834.

White, G. & Powell, W. 1997*a* Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. *Molec. Ecol*. **6**, 851–860.

White, G. & Powell, W. 1997*b* Cross-species amplification of SSR loci in the Meliaceae family. *Molec. Ecol*. **6**, 1185–1197.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.